

# Emergence of Bursts and Communities in Evolving Weighted Networks

Hang-Hyun Jo, Raj Kumar Pan, and Kimmo Kaski

*BECS, Aalto University School of Science, P.O. Box 12200, FI-00076*

Understanding the patterns of human dynamics and social interaction, and the way they lead to the formation of an organized and functional society are important issues especially for techno-social development. Addressing these issues of social networks has recently become possible through large scale data analysis of e.g. mobile phone call records, which has revealed the existence of modular or community structure with many links between nodes of the same community and relatively few links between nodes of different communities. The weights of links, e.g. the number of calls between two users, and the network topology are found correlated such that intra-community links are stronger compared to the weak inter-community links. This feature is known as Granovetter's "The strength of weak ties" hypothesis. In addition to this inhomogeneous community structure, the temporal patterns of human dynamics turn out to be inhomogeneous or bursty, characterized by the heavy tailed distribution of time interval between two consecutive events, i.e. inter-event time. In this paper, we study how the community structure and the bursty dynamics emerge together in a simple evolving weighted network model. The principal mechanisms behind these patterns are social interaction by cyclic closure, i.e. links to friends of friends and the focal closure, i.e. links to individuals sharing similar attributes or interests, and human dynamics by task handling process. These three mechanisms have been implemented as a network model with local attachment, global attachment, and priority-based queuing processes. By comprehensive numerical simulations we show that the interplay of these mechanisms leads to the emergence of heavy tailed inter-event time distribution and the evolution of Granovetter-type community structure. Moreover, the numerical results are found to be in qualitative agreement with empirical analysis results from mobile phone call dataset.

## I. INTRODUCTION

Human dynamics and social interaction patterns have been a subject of intensive study in many different fields ranging from sociology and economics to computer science and statistical physics constituting what is nowadays called network science [1–4]. Partially due to the fact that huge amounts of various kinds of digital data on human dynamics have become available, explorative and quantitative analysis of these kinds of data has enabled us to have unprecedented insight into the structure and dynamics of behavioral, social, and even societal patterns. Examples of such data consist of email correspondence [5, 6], mobile phone call (MPC) and Short Message (SM) communication [7–9], online social network services [10, 11], and scientific collaboration [12].

The interaction structure among individuals in such large scale social data has been investigated by applying the concepts and methods of complex networks where individuals and their relationships represent nodes and links, respectively [13–15]. In many real networks, the link is characterized by a weight corresponding to the strength or closeness of social relationship [15, 16], which in the case of MPC can be described by the aggregate number of calls between two individuals [7, 17]. It has turned out that social networks are inhomogeneous and they can be characterized by modular or community structure [18]: The whole network is composed of separate communities connected by bridges, i.e. there are more and stronger links within communities than between communities, in accordance with Granovetter's "The strength of weak ties" hypothesis [19], corroborated

later in [7, 17]. This weight-topology coupling was successfully reproduced in the model of weighted networks driven by the cyclic and the focal closure processes [20]. Here the cyclic closure process refers to the link formation with one's next nearest neighbors, i.e. the link formation with friends of friends. The focal closure refers to the attribute-related link formation independently of the local connectivity [21]. It has been shown that these simple processes can lead to the emergence of complex weight-topology coupling, where the inhomogeneity of weights is a crucial factor for the emergence of communities.

In addition to the inhomogeneous community structure of social networks, the temporal patterns of human dynamics are inhomogeneous or bursty [6, 22, 23]. The bursts of rapidly occurring events of activity are separated by long periods of inactivity. The bursty dynamics is characterized by the heavy tailed distribution of inter-event times  $\tau$ , defined as the time interval between consecutive events, shows a power-law decay as  $P(\tau) \sim \tau^{-\alpha}$  with  $\alpha \approx 0.7$  or 1 for the MPC [23] or for the email [6], respectively. Two mechanisms for the origin of burstiness have been suggested: a) inhomogeneity due to the human circadian and weekly activity patterns [24, 25] and b) inhomogeneity rooted in the human task execution [6, 22]. Although such dynamic inhomogeneity is obviously affected by the circadian and weekly patterns, it was claimed that the burstiness turns out to be robust with respect to the removal of circadian and weekly patterns from the time series of MPC and SM activities [26]. Here we will concentrate on considering the dynamic inhomogeneities other than those due to circadian and weekly patterns, namely due to those related to individual behavior.

In relation to the inhomogeneity of human task execution, several priority-based queuing models have been studied [6, 22, 27–30]. Each individual is assumed to have a task list of finite size and select one of tasks under the selection protocol, such as selecting the task with the largest priority. Most of these models focus on the waiting time of task, which is defined as the time interval between the arrival time and the execution time of the task. However, in some cases, since the arrival times of tasks to the queue are not given, the waiting times cannot be empirically measured and thus cannot be directly compared to the empirical inter-event times. Furthermore, in spite of studying the communication patterns, such as the email correspondence, the interaction between individuals has not been properly considered in the models [6, 22]. Some interactive models defined on networks assume that the underlying networks are binary and fixed [27–30]. However, in reality both the topology and the weights of social networks co-evolve according to the individual task executions as well as to social interaction by cyclic and focal closure mechanisms.

Both the structural inhomogeneity of social interaction and the dynamical inhomogeneity of human individual behavior affect the dynamical processes taking place on evolving social networks. For example it has been shown that the Granovetter-type weight-topology coupling slows down information spreading [7]. By using the analogy between link weight and information-bandwidth information turns out to spread fast and to get trapped within communities due to the internal strong links (broad bandwidth) and the weak links (narrow bandwidth) between communities, respectively. In addition to the effect of Granovetter-type community structure on information spreading individual bursty behavior also plays a crucial role in social dynamics. The long inactive periods represented by large inter-event times, inhibit the information spreading compared to the randomized null model, while the bursty periods of short inter-event times do not necessarily enhance the spreading [23]. Thus both the weight-topology coupling and the individual bursty dynamics should be taken into account and implement to a model in order to better understand the dynamics in the evolving social networks.

The observation of Granovetter-type community structure and individual bursty dynamics calls for integrating both structural and dynamical inhomogeneities into single framework or model in order to better understand the social dynamics with the smallest set of parameters. Although there are some approaches in integrating these structural and dynamical properties, the bursty nature of human behavior has been inherently assumed in these models [31, 32]. Instead, we are interested in the emergence of burstiness from the intuitive and natural model rules while at the same time generating the Granovetter-type community structure. In order to investigate the basic mechanisms responsible for various empirical observations, we incorporate the task handling process to the weighted network formation studied by Kumpula *et*

*al.* [20]. In our model the weight assigned to a link is interpreted as the aggregate number of events on that link. Driven by both the cyclic and focal closure mechanisms a link is created by the first event occurring between individuals. Once created, the link is maintained by a series of events on that link, and finally removed by accidental memory loss of the individual. Each individual may initiate events or respond to those initiated by others, depending on the protocols determining the selection and execution of tasks given to individuals.

Our model can be called co-evolutionary, in the sense that the task handling process of individuals affects the network evolution while the network structure constrains the individual behaviors. One of the typical issues in the co-evolutionary networks is that the timescale of network evolution competes with that of the dynamical process on the network [33, 34]. In social dynamics the timescale for social relationship updates (a few weeks or months) is much larger than communication dynamics taking place on daily or hourly basis. In our case, since the events are the building blocks of the structure and the dynamics simultaneously, the relevant timescales are not explicitly controlled but emerged from the simple and intuitive rules of our model. In this paper, we show that by using the models with a few control parameters one can obtain the Granovetter-type community structure and also observe the emergence of bursty dynamics characterized by the heavy tailed inter-event time distribution.

This paper is organized such that we first introduce our two kinds of co-evolutionary models, the Triad-Interaction-enhanced model and the Process-Equalized model. Then we present the results for these models and discuss them in comparison with the empirical analysis results followed by the conclusions on the findings in the paper.

## II. METHODS

In our model we assume that the network evolves by means of link creation, link maintenance, and link deletion. Once a link between two stranger nodes is created by either the cyclic or the focal closure mechanisms, it is maintained by a series of events on that link, which we call the neighboring interaction (NI), or it is deleted by random memory loss. The focal closure mechanism is implemented by the random pairing of nodes, which is called global attachment or GA process. The cyclic closure mechanism is realized when a node interacts with its next nearest neighbor, which is called local attachment or LA process. While the GA process involves dyad interaction, the LA process is mediated by the third node, implying triad interaction. The NI process between neighboring nodes can happen directly, i.e. as dyad NI, or can be mediated by their common neighbor, i.e. as triad NI. Let us assume that only the event like the peer-to-peer phone call is considered. Then we can implement the triad interaction by splitting it into dyad interactions

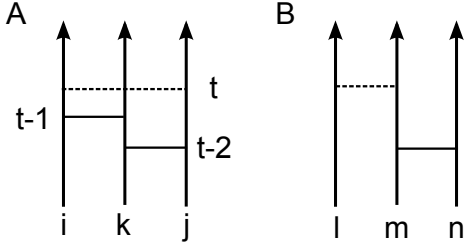


FIG. 1: **Local and global attachments of the model.** Vertical lines represent the time lines of users. Horizontal solid and dashed lines represent events occurred and to be occurred between nodes, respectively. The number of events on a link defines the weight of the link. **A.** Local Attachment (LA): The node  $i$  has a chance to interact with node  $j$  at time step  $t$  only when there exists a temporal path connecting  $i$  and  $j$  through their common neighbor  $k$  within time window  $[t-2, t-1]$ . **B.** Global Attachment (GA): The isolated node  $l$  has a chance to interact with a randomly chosen node  $m$ .

such that a node  $i$  has a chance to interact with  $j$  at time step  $t$  only when both  $i$  and  $j$  have interacted with the third node  $k$  recently, no more than, say, 2 time steps before, see fig. 1. In the following we propose two kinds of models. In the first kind the triad interaction takes place prior to the dyad interaction. We call this as Triad-Interaction-enhanced model (TI model in short). The TI model is a direct extension of the weighted network model by Kumpula *et al.* [20], where the dyad NI process is analogous with Barabási's task execution model [6]. In the second kind all the three processes (LA, GA, and NI) are considered equally and the framework of interacting and non-interacting tasks is adopted from [27], as the variant of Barabási's task execution model. We call this as Process-Equalized model (PE model in short).

Now let us consider an undirected weighted network with  $N$  nodes. A weight of a link between nodes  $i$  and  $j$ , denoted by  $w_{ij}$ , can be interpreted as the aggregate number of events between them. The number of neighbors of node  $i$  is defined as the degree  $k_i$ . The time step of the most recent event between node  $i$  and node  $j$  is denoted by  $t_{ij}$ . Initially all nodes are set to be isolated, i.e. the initial network is without links.

### A. Triad-Interaction-enhanced model

In the TI model the dynamics at each time step  $t$  consists of the following three stages:

**1) Triad Interaction (LA and triad NI):** For each pair of nodes  $i$  and  $j$  satisfying  $\{t_{ik}, t_{jk}\} = \{t-2, t-1\}$  with a third node  $k$ , we check whether  $i$  and  $j$  are connected. If they are connected, an event between  $i$  and  $j$  occurs, i.e.  $w_{ij} \rightarrow w_{ij} + 1$ , corresponding to the triad NI process. Otherwise, the event between  $i$  and  $j$  occurred with probability  $p_{LA}$  leads to  $w_{ij} = 1$ , implying a link creation by the LA process. These LA and triad NI processes are responsible for the community formation

and weight reinforcement, respectively.

**2) Dyad Interaction (GA and dyad NI):** Every node not involved in the previous stage selects a target node to make an event. If isolated, the node selects the target node from the whole population at random, preparing for the GA process. If non-isolated, the node selects the target node either from the whole population or from its neighbors with probabilities  $p_{GA}$  or  $1 - p_{GA}$ , respectively. In other words, all nodes are free to find new neighbors while the non-isolated nodes are also responsible for maintaining links to the existing neighbors, the degree of which is controlled by  $p_{GA}$ . In the case of selecting the target from its neighbors, preparing for the dyad NI process, the probability of the node  $i$  selecting its neighboring node  $j$  is proportional to the weight between them,  $w_{ij}$ . Thus there is preference for the strong links. Targeting  $j$  by  $i$  is denoted by  $i \rightarrow j$ . The analogy between the target selection from the population or from the neighbors and the task selection from the task list will be discussed later.

The nodes having selected their targets make events with targets in a random order only when both the node and its target are not yet involved in any other event at this time step. If the node  $i$  and its target  $j$  were not connected, the event leads to a link creation between them, i.e. the realization of the GA process. Otherwise, the event between them results in  $w_{ij} \rightarrow w_{ij} + 1$ , implying the dyad NI process.

**3) Memory Loss:** With probability  $p_{ML}$ , each node,  $i$ , becomes isolated and a stranger to all its neighbors  $j$  as  $w_{ij} = 0$ . This completes the time step  $t$ .

Through all the above stages it has been assumed that the target has no choice to reject the event initiated by some other node. We term this the OR protocol [28] in a sense that it is enough for at least one of two nodes to initiate and make an event between them. Hence we call this version as the TI-OR model. Alternatively we can assume that an event can occur only in the reciprocal case, i.e.  $i \rightarrow j$  and  $j \rightarrow i$ , which implies the AND protocol. It should be noted that for example a mobile phone user can reject a call from his/her friend by some reason. Here we will consider an TI-AND model, where the AND protocol is applied only to the dyad NI process.

### B. Process-Equalized model

In the TI model, since the triad interaction is executed prior to the dyad interaction, one can not control the intensity of the triad interaction. Therefore we have devised the PE model where we consider the triad interaction on the equal footing with the dyad interactions, i.e. the LA, GA, and NI processes are equally considered. In this case we incorporate the task execution process with interacting and non-interacting tasks [27], as described next.

Each node has the task list with one interacting task and one non-interacting task, denoted by  $I$ -task and  $O$ -

task, respectively. The  $I$ -task represents the task requiring simultaneous interaction of two nodes, such as a phone call by a caller to a receiver, while the  $O$ -task represents some other task not requiring the simultaneity such as shopping, watching TV, etc. We count the inter-event times only for  $I$ -tasks, which settles down the issue of realistically interpreting the waiting time, as mentioned in [27]. The priorities of tasks are randomly drawn from the uniform distribution.

In this model the dynamics takes place such that at each time step  $t$ , every node selected in a random order goes through the stages 1) and 2). Then the stage 3) is performed:

**1) Task and Target Selection:** The node selects the task with larger priority. Only when it is  $I$ -task, this node, which we call a root node  $i$ , selects its target node either

- from the whole population with probability  $p_{GA}$ , i.e. the GA process, or
- from its next nearest neighbors with probability  $p_{LA}$ , i.e. the LA process, or
- from its neighbors with probability  $1 - p_{GA} - p_{LA}$ , i.e. the NI process.

For the LA process, the next nearest neighbor of the root node is defined as the node  $j$  satisfying  $\{t_{ik}, t_{jk}\} = \{t-2, t-1\}$  with another node  $k$ . If the number of next nearest neighbors is more than 1, one of them is selected at random. For the NI process, the probability to target one of the root node's neighbors  $j$  is proportional to the weight  $w_{ij}$ , as in the TI model.

**2) Task Execution:** Only when the target node has not been involved in any event at this time, the event between the root node and the target node occurs, implying that the OR protocol is used. After this execution the priority of the  $I$ -task for the root node is replaced by the new random number while  $j$ 's task list is not updated, implying that the target node did not execute its  $I$ -task but simply responded to the root node.

**3) Memory Loss:** Each node becomes isolated with probability  $p_{ML}$ , by which the time step  $t$  is completed.

### C. Definitions of network properties

We calculate various network properties for the numerically obtained networks. Given the weight distribution  $P(w)$ , the cumulative weight distribution is defined by

$$P_c(w) \equiv \int_w^\infty P(w') dw'. \quad (1)$$

For each non-isolated node  $i$ , the number of next nearest neighbors, the individual clustering coefficient, and the

strength are defined by

$$k_{nn,i} \equiv \frac{1}{k_i} \sum_{j \in \Lambda_i} k_j, \quad (2)$$

$$c_i \equiv \frac{2e_i}{k_i(k_i - 1)}, \quad (3)$$

$$s_i \equiv \sum_{j \in \Lambda_i} w_{ij}, \quad (4)$$

respectively. Here  $\Lambda_i$  denotes the set of neighbors and  $e_i$  denotes the number of links among the node  $i$ 's neighbors. The averages of the above quantities over the nodes with the same degree  $k$  define the average number of next nearest neighbors  $k_{nn}(k)$ , the local clustering coefficient  $c(k)$ , and the average strength  $s(k)$ , respectively. In addition, to test the Granovetter-type community structure the overlap is defined for each link connecting nodes  $i$  and  $j$  as follows:

$$O_{ij} \equiv \frac{|\Lambda_i \cap \Lambda_j|}{|\Lambda_i \cup \Lambda_j|}, \quad (5)$$

i.e. the fraction of the common neighbors over all neighbors of  $i$  and  $j$ . The average over the links with the same weight  $w$  defines the average overlap  $O(w)$ . For the dynamics we measure the inter-event time distributions  $P(\tau)$ .

## III. RESULTS AND DISCUSSION

The empirical analysis of mobile phone call data from a single operator in one European country for the first four months in 2007 [17, 23] shows that  $c(k) \sim k^{-\delta_c}$  with  $\delta_c \approx 1$ ,  $s(k) \sim k^{\delta_s}$  with  $\delta_s \approx 1$ . It also shows an increasing behavior of  $k_{nn}(k)$ , implying the assortativity, and an increasing behavior of  $O(w)$  with slight decrease for very large  $w$  values, where the increasing part implies the Granovetter-type community structure. Moreover, it was found that  $P(\tau) \sim \tau^{-\alpha}$  with  $\alpha \approx 0.7$ . In addition the average degree  $\langle k \rangle$  turned out to be around 3.0 and the average clustering coefficient  $\langle c \rangle$  around 0.13 when the new year's day of 2007 is excluded. It should be noted that the average degree of the mobile phone call network extracted from the single operator dataset might be underestimated compared to the full mobile phone call network composed by many operators. Therefore, we assume that the overall average degree of the whole social network is larger than 3, i.e. around 10. In this paper we consider the results to be relevant and comparable with reality, when  $\langle k \rangle \approx 10$ .

For the numerical simulations of the models described above we set the initial values as  $N = 5 \times 10^4$  and  $p_{ML} = 10^{-3}$  for all the cases considered. The simulations of these models are found to become stationary at about  $t = 3 \times 10^3$ , after which the numerical results are collected for  $5 \times 10^4$  time steps.

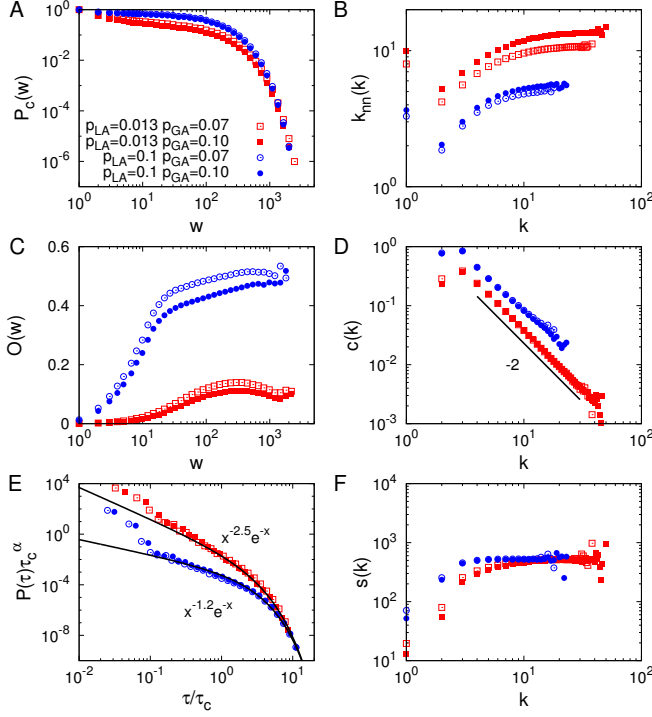


FIG. 2: **TI-OR model.** **A.** The cumulative weight distribution  $P_c(w)$ . **B.** The average number of next nearest neighbors  $k_{nn}(k)$ . **C.** The average overlap  $O(w)$ . **D.** The local clustering coefficient  $c(k)$ . **E.** The inter-event time distribution  $P(\tau)$ . **F.** The average strength  $s(k)$ . Results are averaged over 50 realizations for networks with  $N = 5 \times 10^4$  and  $p_{ML} = 10^{-3}$ . We obtain  $\langle k \rangle \approx 10.1$  and  $\langle c \rangle \approx 0.08$  for  $p_{LA} = 0.013$  and  $p_{GA} = 0.1$ . The cases with  $p_{LA} = 0.1$  and/or with  $p_{GA} = 0.07$  are also plotted for comparison.

### A. Triad-Interaction-enhanced model

For both the TI-OR and TI-AND models, we find that the cumulative weight distributions are broad but do not follow power-law behavior for the various values of  $p_{LA}$  and  $p_{GA}$  as in the empirical analysis, see figs. 2 and 3. The similar behavior is found for degree and strength distributions (not shown). It turns out that as the empirical results, the networks are assortative and have the Granovetter-type community structure, characterized by the increasing behavior of  $k_{nn}(k)$  for  $k \geq 2$  and  $O(w)$ , respectively. Here most nodes with  $k = 1$  are supposed to be connected to randomly chosen nodes by the GA process, implying that  $k_{nn}(1) \approx \langle k \rangle$ . The sample networks shown in fig. 5A-D also confirm the emergence of Granovetter-type community structure, such that the communities of internal strong links are connected by weak links. In addition, for the TI-OR model with  $p_{LA} = 0.1$  we observe a slightly decreasing behavior of  $O(w)$  for large  $w$  values, implying the existence of smaller but stronger communities. The decreasing behavior of the overlap was observed in the empirical analysis but not in the previous model studies [20].

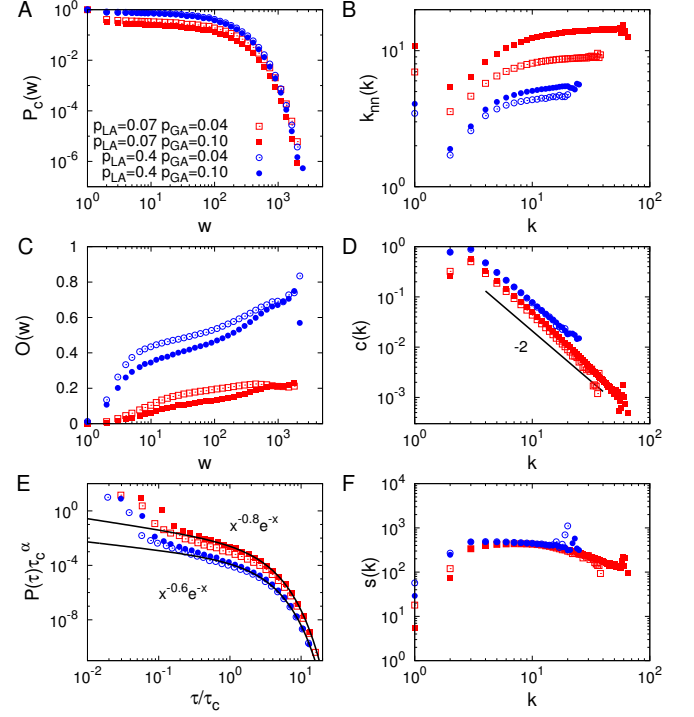


FIG. 3: **TI-AND model.** **A.** The cumulative weight distribution  $P_c(w)$ . **B.** The average number of next nearest neighbors  $k_{nn}(k)$ . **C.** The average overlap  $O(w)$ . **D.** The local clustering coefficient  $c(k)$ . **E.** The inter-event time distribution  $P(\tau)$ . **F.** The average strength  $s(k)$ . Results are averaged over 50 realizations for networks with  $N = 5 \times 10^4$  and  $p_{ML} = 10^{-3}$ . We obtain  $\langle k \rangle \approx 9.6$  and  $\langle c \rangle \approx 0.13$  for  $p_{LA} = 0.07$  and  $p_{GA} = 0.1$ . The cases with  $p_{LA} = 0.4$  and/or with  $p_{GA} = 0.04$  are also plotted for comparison.

Based on the above observations it seems that a node is a member of a few strong triangles and connected to some other nodes outside its own triangles. This explains our finding of  $\delta_c \approx 2$ , different from the empirical result. It is because if the degree of a node increases mainly by means of the GA process, the number of links between neighbors remains while the number of all possible links grows as  $k^2$ , resulting in  $c(k) \sim k^{-2}$ . We also find  $\delta_s \approx 0$  differently from the empirical value, which we will discuss later in relation to the dynamics.

In order to confirm the Granovetter-type community structure of networks, we perform the link percolation analysis. If links within communities are strong whereas links between them are weak as found in the empirical studies [7, 17], the network should disintegrate faster when the weak links are removed first than when the strong links are removed first. Note that as shown in fig. 5 the links with weight 1 form an apparently random network as backgrounds for the community structure. Thus we apply the link percolation to the giant components of networks without links with  $w = 1$  and denote its size by  $N'$ . By removing links in an ascending or descending order of weights, we measure the remaining fraction

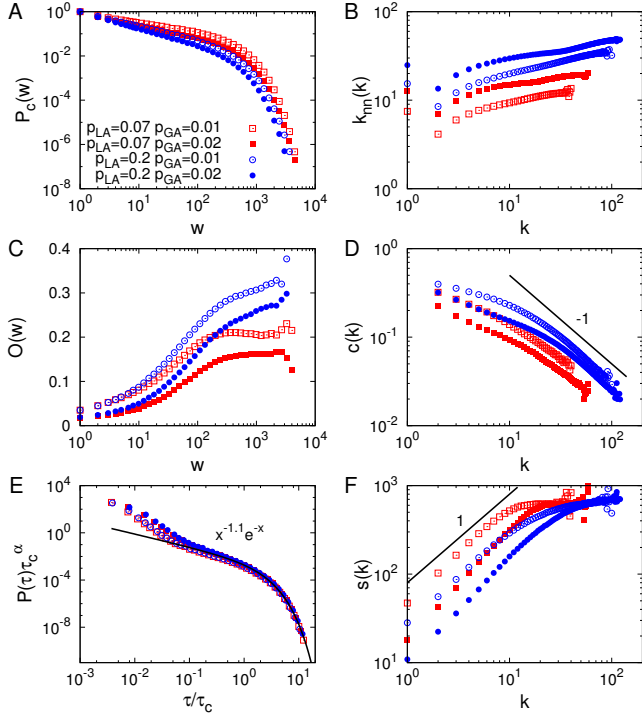


FIG. 4: **PE model.** **A.** The cumulative weight distribution  $P_c(w)$ . **B.** The average number of next nearest neighbors  $k_{nn}(k)$ . **C.** The average overlap  $O(w)$ . **D.** The local clustering coefficient  $c(k)$ . **E.** The inter-event time distribution  $P(\tau)$ . **F.** The average strength  $s(k)$ . Results are averaged over 50 realizations for networks with  $N = 5 \times 10^4$  and  $p_{ML} = 10^{-3}$ . We obtain  $\langle k \rangle \approx 9.9$  and  $\langle c \rangle \approx 0.11$  for  $p_{LA} = 0.07$  and  $p_{GA} = 0.02$ . The cases with  $p_{LA} = 0.2$  and/or with  $p_{GA} = 0.01$  are also plotted for comparison.

of the giant component  $R_{GC}$ , the susceptibility  $\chi$ , and the average clustering  $\langle c \rangle$  as a function of the fraction of removed links,  $f$ . Here the susceptibility is defined as  $\chi = \sum n_s s^2 / N'$ , where  $n_s$  denotes the number of clusters with size  $s$  and the giant component is excluded from the summation. For the weak-link-first-removal cases we find the sudden disintegration of networks at the finite value of  $f$ , i.e.  $f_c = 0.62$  (0.21) for TI-OR (TI-AND) model. When the strong links are removed first, there is an apparent transition at the larger value of  $f_c = 0.87$  (0.81) for TI-OR (TI-AND) model as shown in fig. 6A and C. For the weak-link-first-removal cases the values of  $f$  maximizing  $\langle c \rangle$ , denoted by  $f_{max}$ , are quite close to those of  $f_c$ . When using the overlap instead of the weight for the link percolation, almost the same behavior is observed in fig. 6B and D because  $O(w)$  turns out to be the monotonically increasing function of  $w$  in our model.

For the temporal dynamics the inter-event time distributions are characterized by the power-law with an exponential cutoff, i.e.  $P(\tau) \sim \tau^{-\alpha} \exp(-\tau/\tau_c)$ , where the scaling regimes span over about one decade, see figs. 2E and 3E. In case of TI-OR model,  $\alpha \approx 2.5$  or 1.2 when  $p_{LA} = 0.013$  or 0.1, respectively. In the case of TI-AND

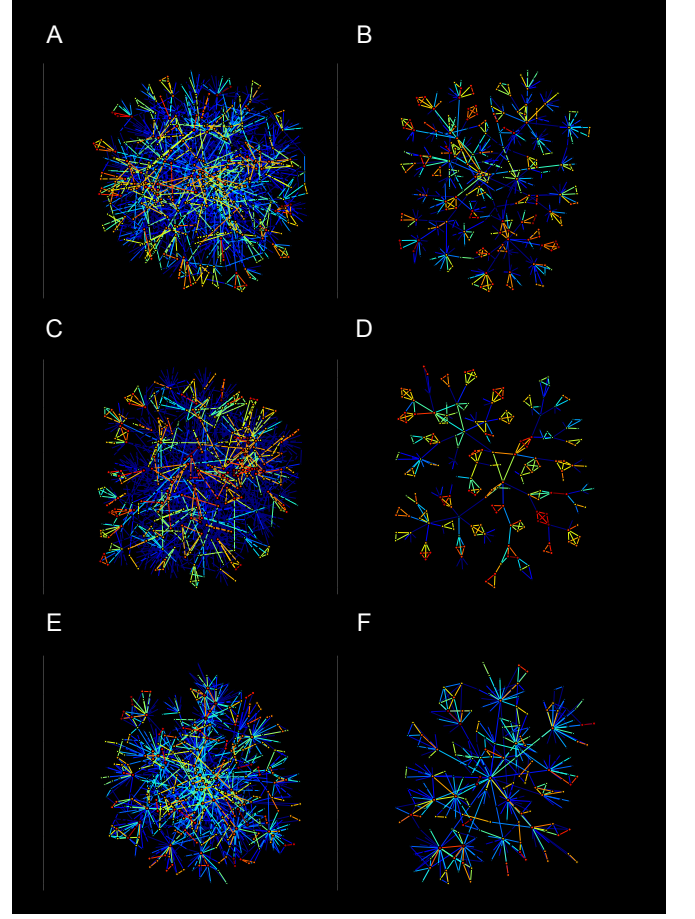


FIG. 5: **Snowball samples [35] of networks for TI-OR model (A, B), for TI-AND model (C, D) and for PE model (E, F).** For each model, we plot the original sample starting from a random node (A, C, E) and the one without the links with  $w = 1$  (B, D, F) for the clear visualization. The color of links ranges from blue for weak links through yellow for intermediate links to red for strong links.

model, when  $p_{LA} = 0.07$  or 0.4, we find  $\alpha \approx 0.8$  or 0.6, respectively, both of which are close to the empirical value 0.7 of MPC dataset within error bars. In all cases, the values of  $\alpha$  are smaller for larger values of  $p_{LA}$  but are barely affected by the value of  $p_{GA}$ . The values of  $\tau_c$  turn out to be larger for larger values of  $p_{LA}$  and for smaller values of  $p_{GA}$ . The maximum value of  $\tau_c$  is around 50.

To figure out what are the possible underlying mechanism for these findings, we first identify the triangular chain interaction (TCI) among three neighboring nodes, say  $i$ ,  $j$ , and  $k$ : Both the event between nodes  $i$  and  $j$  at time step  $t - 2$  and the event between nodes  $j$  and  $k$  at time  $t - 1$  lead to an event between nodes  $i$  and  $k$  at time  $t$ , again leading to another event between nodes  $i$  and  $j$  at time  $t + 1$  and so on, unless interrupted either by the events from/to nodes outside the triangle or by a random memory loss of nodes in the triangle. Since the TCI is exclusive due to the priority of the triad interaction including the LA process, the LA process enhanced



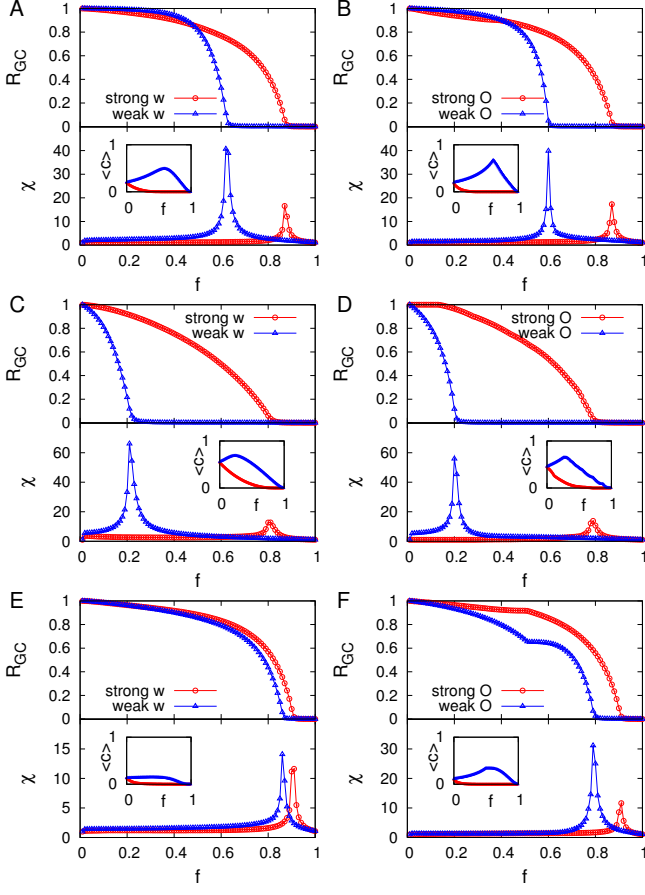


FIG. 6: **Link percolation analysis for TI-OR model (A, B), for TI-AND model (C, D), and for PE model (E, F).** A, C, E. The weight is used as the link strength. B, D, F. The overlap is used as the link strength. For each panel, we calculate the fraction of giant component  $R_{GC}$ , susceptibility  $\chi$ , and clustering coefficient  $\langle c \rangle$  (inset) as a function of the fraction of removed links,  $f$ . Results are averaged over 50 realizations for networks originally with  $\langle k \rangle \approx 10$  for each model.

by the large value of  $p_{LA}$  inhibits the interruption by the events from/to nodes outside the triangle, including the GA process, and thus making the community structure more compact in turn resulting in a smaller average degree. In case of TI-OR model with  $p_{GA} = 0.1$ ,  $\langle k \rangle \approx 10.1$  or 4.2 for  $p_{LA} = 0.013$  or 0.1, respectively. While the compact community structure enhances the TCI again, explaining the observed peaks of  $P(\tau)$  at  $\tau = 1$  and 2, it can also make some neighbors of the TCI nodes wait for long time to interact with the TCI nodes. Hence, the larger value of  $p_{LA}$  gives rise to larger fluctuation in the inter-event times, implying a smaller value of the power-law exponent  $\alpha$  and a larger value of the cutoff  $\tau_c$ , as observed. Based on this argument, the effect of  $p_{LA}$  dominates over that of  $p_{GA}$ , so that the value of  $p_{GA}$  barely affects the scaling of inter-event time distributions but it controls the value of  $\tau_c$ . The larger value of  $p_{GA}$  al-

lows nodes to choose a random target and thus interrupt the inter-event times of targets more frequently, leading to a smaller value of  $\tau_c$ . The numerical results in the case of the TI-AND model can be explained by the same arguments, except for the observed values of  $\alpha$  less than those found in the case of the TI-OR model. Note that in general the AND protocol inhibits the possibility of events.

The heavy tailed distribution of inter-event times, i.e. bursty dynamics, was not expected but it emerged from the model. Analogously with the task execution model suggested by Barabási [6], the dyad NI process can be interpreted such that a node  $i$  has the task list with size  $k_i$  and it selects one of neighbors (tasks)  $j$  with probability proportional to the priority of the task, i.e. the weight  $w_{ij}$  in our model. The degree  $k_i$  also varies depending on the link creation and deletion processes. A node having been isolated by the memory loss tries to interact with strangers. Once being connected to some other node by the GA process, its degree increases partly by means of the LA process but it will not diverge. The degree mostly fluctuates and sometimes remains unchanged for long periods of time. And the node finally becomes isolated again by the memory loss. Thus, the whole life-cycle of a node is assumed to consist of two types of periods, i.e. one with fixed-size and the other with variable-size task list. The periods of fixed-size task list, i.e. fixed degrees, are up to several hundred time steps, which are much larger than the observed  $\tau_c$ . This implies the natural separation of timescales between network change and dynamics on the network, which is consistent with everyday experience of mobile phone usage. Due to the timescale separation the inter-event time distribution for the whole period can be represented by the superposition of those for fixed-size period and for variable-size period. Thus, to understand the effect of size variability on the scaling behavior of bursty dynamics, we refer to the previous works studied in the different kinds of models, such as by Vázquez *et al.* [22]. When the task list has a variable (fixed) size in the Barabási model, the power-law exponent for the waiting time distribution turns out to be 3/2 (2). According to the argument that the distribution of the inter-event times derived from the waiting times has the same power-law exponent as that of the waiting times, one can expect the similar values of exponent from our model. However, this is not the case with our model, so we leave this for the more rigorous analysis in the future.

Finally, the apparent overall independence of the average strength  $s(k)$  on  $k$  for large values of  $k$  is attributed to the fact that once the node is a member of the TCI, its activity becomes effectively independent of its degree due to the exclusive property of TCI. We observe even the decreasing behaviors of  $s(k)$  for the larger  $k$  values in the TI-AND model, i.e. the AND protocol based interaction with too many neighbors can make nodes failing to interact with any neighbors.

## B. Process-Equalized model

The TI models show the expected behaviors of Granovetter-type community structure and the heavy tailed inter-event time distribution but they do not yield the expected behavior of the local clustering coefficient and average strength of the nodes. This is mainly due to too strong effect of the triad interaction and that is why we need to consider the PE model for modeling improvement and comparison with empirical results.

With the PE model we find that the cumulative weight distributions  $P_c(w)$  are broad, that the overlap  $O(w)$  increases with  $w$ , i.e. showing Granovetter-type community structure, that the average number of next nearest neighbor  $k_{nn}(k)$  increases with  $k$ , i.e. showing the network being assortativity for  $k \geq 2$ , and that  $c(k) \sim k^{-\delta_c}$  with  $\delta_c \approx 1$  and  $s(k) \sim k^{\delta_s}$  with  $\delta_s \approx 1$ , as shown in fig. 4. All these results are consistent with the empirical analysis on real data. Based on the sample networks in fig. 5E and F, it is evident that the TCI becomes weaker and less exclusive than in the case of the above TI models. Therefore, as the degree of a node increases, the neighbors of that node have the increasing chance to interact with each other, resulting in  $c(k) \sim k^{-1}$ .

In fig. 6E and F we show the results of the link percolation analysis, done to confirm Granovetter-type community structure. We find that when the weak links are removed first, the percolation transition occurs at  $f_c = 0.86$ . On the other hand when the strong links are removed first, a transition is observed at  $f_c = 0.91$ , implying that the strong links play the role of bridges between communities. This is also evident in the sample networks in fig. 5E and F. The curve of the average clustering coefficient  $\langle c \rangle$  turns out to be flat for a wide range of  $f$  values. Similar behaviors are also observed when the overlap is used instead of the weight in the link percolation analysis. For the weak-link-first-removal we find  $f_c = 0.79$  and  $f_{max} = 0.55$ , where yet another kink in the curve of  $R_{GC}$  is observed. This implies that the network goes through two abrupt changes, first at  $f_{max}$  and then at  $f_c$ .

Here we also observe the heavy tailed distributions of inter-event times with exponential cutoffs following a power law behavior with the exponent of  $\alpha \approx 1.1$ . The task execution model for each node would result in  $\alpha = 1$  as in the case of Barabási's queuing model if only initiating the  $I$ -tasks are counted as the relevant events and if the neighbors of the node always respond to that node. However, the nodes are supposed to interact with each other such that by initiating  $I$ -tasks some root nodes can interrupt the inactive periods of their target nodes, which in general decreases the inter-event times. On the other hand, if the target is already involved in another event so that the trials by the root nodes fail, the inter-event times of corresponding root nodes would increase up to the points of next successful events occurring. The observed value of  $\alpha \approx 1.1$  indicates that any of the mentioned factors did not affect much the scaling behavior of

the distributions. The values of  $\tau_c$  are largely or barely affected by the value of  $p_{GA}$  or  $p_{LA}$ , respectively, in an anti-correlated way. The maximum value of  $\tau_c$  is around 270.

The observation of the average strength  $s(k) \sim k$  behavior can be explained by considering the dynamics where the OR protocol is adopted. In this case the nodes with many neighbors might receive more calls from their neighbors than those with few neighbors do, while the chance to make calls is the same for any node.

## C. Conclusions

We have studied the emergence of Granovetter-type community structure, characterized by the increasing behavior of overlap as a function of the link weight, and the heavy tailed inter-event time distributions, i.e. bursty dynamics, in a single framework of simultaneously evolving weighted network model. By incorporating simple and intuitive task execution models for human dynamics into the weighted network model reproducing the Granovetter-type community structure of social systems, we successfully observe the qualitatively same behaviors as observed in the empirical networks based on the mobile phone call (MPC) dataset. In addition, we have found that the exclusive triangular chain interaction (TCI) identified in the TI models plays the central role both in community structure formation and bursty dynamics. For the existence of TCI we have the evidence from the empirical study on the dynamic motifs of MPC communication [36]. The numerical results from TI-OR and TI-AND models are qualitatively the same except for the power-law exponent  $\alpha$  of inter-event time distributions. The values of  $\alpha$  from TI-AND model turn out to be closer to the empirical value 0.7 for the MPC, implying that the AND protocol is necessary to properly model the MPC communication. Furthermore, in the PE model, by relaxing the exclusive property of TCI to some extent we could obtain more realistic results at least for the network structure. The scaling behavior of inter-event time distributions seems to be mainly affected by the incorporated framework of interacting and non-interacting tasks, which should be made clear in the future.

Finally we believe that building simple empirical-observation-based models, like our TI- and PE-models, by incorporating the process of human task execution by priority-based queuing with the basic processes of friendship-network formation by cyclic and focal closure mechanisms enable us to better understand the underlying mechanisms of real co-evolutionary networks. Furthermore, these models enable us to explore the social dynamics in these networks as done differently by Karsai *et al.* [23] with the susceptible-infected (SI) dynamics for the mobile phone call communication. Moreover, the scaling properties and finite-size scaling of real networks are usually not so informative but can be considered and made more informative by means of simple but still quite



realistic models, where one can control the system size and other parameters as well.

Center of Excellence program 2006-2011, project no. 129670 (RKP, KK) are gratefully acknowledged.

### Acknowledgments

Financial support by Aalto University postdoctoral program (HJ), by the Academy of Finland, the Finnish

- 
- [1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*, 1st ed., Structural analysis in the social sciences, 8 (Cambridge University Press, 1994).
  - [2] S. Goyal, *Connections : An Introduction to the Economics of Networks* (Princeton University Press, Princeton, NJ, 2007).
  - [3] M. E. J. Newman, *Networks: An Introduction*, 1st ed. (Oxford University Press, USA, 2010).
  - [4] A.-L. Barabási, *Bursts: The Hidden Pattern Behind Everything We Do* (Dutton Books, 2010).
  - [5] J.-P. Eckmann, E. Moses, and D. Sergi, Proceedings of the National Academy of Sciences of the United States of America **101**, 14333 (2004).
  - [6] A.-L. Barabási, Nature **435**, 207 (2005).
  - [7] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási, Proceedings of the National Academy of Sciences **104**, 7332 (2007).
  - [8] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, Nature **453**, 779 (2008).
  - [9] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber, Proceedings of the National Academy of Sciences **107**, 18803 (2010).
  - [10] H. Kwak, C. Lee, H. Park, and S. Moon, in *Proceedings of the 19th international conference on World wide web, WWW '10* (ACM, New York, NY, USA, 2010) pp. 591–600.
  - [11] J.-P. Onnela and F. Reed-Tsochas, Proceedings of the National Academy of Sciences **107**, 18375 (2010).
  - [12] M. E. J. Newman, Proceedings of the National Academy of Sciences of the United States of America **98**, 404 (2001).
  - [13] R. Albert and A.-L. Barabasi, Reviews of Modern Physics **74**, 47 (2002).
  - [14] M. E. J. Newman, SIAM Review **45**, 167 (2003).
  - [15] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Physics Reports **424**, 175 (2006).
  - [16] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proceedings of the National Academy of Sciences of the United States of America **101**, 3747 (2004).
  - [17] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, New Journal of Physics **9**, 179 (2007).
  - [18] S. Fortunato, Physics Reports **486**, 75 (2010).
  - [19] M. S. Granovetter, The American Journal of Sociology **78**, 1360 (1973).
  - [20] J. M. Kumpula, J. P. Onnela, J. Saramäki, K. Kaski, and J. Kertész, Physical Review Letters **99**, 228701 (2007).
  - [21] G. Kossinets and D. J. Watts, Science **311**, 88 (2006).
  - [22] A. Vázquez, J. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, and A. L. Barabási, Physical Review E **73**, 036127 (2006).
  - [23] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A. L. Barabási, and J. Saramäki, Physical Review E **83**, 025102 (2011).
  - [24] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. Amaral, Proceedings of the National Academy of Sciences of the United States of America **105**, 18153 (2008).
  - [25] R. D. Malmgren, D. B. Stouffer, A. S. L. O. Campanharo, and L. A. Amaral, Science **325**, 1696 (2009).
  - [26] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski, arXiv:1101.0377 .
  - [27] J. Oliveira and A. Vazquez, Physica A: Statistical Mechanics and its Applications **388**, 187 (2009).
  - [28] B. Min, K. I. Goh, and I. M. Kim, Physical Review E **79**, 056110 (2009).
  - [29] W. K. Cho, B. Min, K. I. Goh, and I. M. Kim, Physical Review E **81**, 066109 (2010).
  - [30] B. Min, K. I. Goh, and A. Vazquez, Physical Review E **83**, 036102 (2011).
  - [31] J. Stehlé, A. Barrat, and G. Bianconi, Physical Review E **81**, 035101 (2010).
  - [32] K. Zhao, J. Stehlé, G. Bianconi, and A. Barrat, Physical Review E **83**, 056109 (2011).
  - [33] P. Holme and M. E. J. Newman, Physical Review E **74**, 056108 (2006).
  - [34] G. Iniguez, J. Kertész, K. K. Kaski, and R. A. Barrio, Physical Review E **80**, 066119 (2009).
  - [35] S. H. Lee, P. J. Kim, and H. Jeong, Physical Review E **73**, 016102 (2006).
  - [36] L. Kovanen, *Structure and dynamics of a large-scale complex social network*, Master's thesis, Aalto University, Espoo, Finland (2009).